

Bridging the Gap Between Ethics and Practice

Year: 2020 | Citations: 356 | Authors: B. Shneiderman

Abstract

This article attempts to bridge the gap between widely discussed ethical principles of Human-centered AI (HCAI) and practical steps for effective governance. Since HCAI systems are developed and implemented in multiple organizational structures, I propose 15 recommendations at three levels of governance: team, organization, and industry. The recommendations are intended to increase the reliability, safety, and trustworthiness of HCAI systems: (1) reliable systems based on sound software engineering practices, (2) safety culture through business management strategies, and (3) trustworthy certification by independent oversight. Software engineering practices within teams include audit trails to enable analysis of failures, software engineering workflows, verification and validation testing, bias testing to enhance fairness, and explainable user interfaces. The safety culture within organizations comes from management strategies that include leadership commitment to safety, hiring and training oriented to safety, extensive reporting of failures and near misses, internal review boards for problems and future plans, and alignment with industry standard practices. The trustworthiness certification comes from industry-wide efforts that include government interventions and regulation, accounting firms conducting external audits, insurance companies compensating for failures, non-governmental and civil society organizations advancing design principles, and professional organizations and research institutes developing standards, policies, and novel ideas. The larger goal of effective governance is to limit the dangers and increase the benefits of HCAI to individuals, organizations, and society.